FINAL DRAFT INTERNATIONAL STANDARD

ISO/IEC 11179-1 Specification and standardization of data elements - Part 1: Framework

Contents

Foreword

ISO (the International Organization for Standardization) and IEC (the International Electrotechnical Commission) form the specialized system for worldwide standardization. National bodies that are members of ISO or IEC participate in the development of International Standards through technical committees established by the respective organization to deal with particular fields of technical activity. ISO and IEC technical committees collaborate in fields of mutual interest. Other international organizations, governmental or non-governmental, in liaison with ISO and IEC, also take part in the work.

In the field of information technology, ISO and IEC have established a joint technical committee, ISO/IEC JCT1. Draft International Standards adopted by the joint technical committee are circulated to national bodies for voting. Publication as an International Standard requires approval of at least 75% of the national bodies casting a vote.

Introduction

Background

Humans are aware of anything that exists in the natural world through its properties. Data represents the properties of these things. Specification of data elements, the basic units of data, involves documenting relevant characteristics of each data element to ensure its representation of the natural world item is consistent and accurate. Data that has been carefully specified and standardized greatly enhances its usefulness and shareability across systems and environments. Sharing data involves the ability to locate desired data, retrieve the data, and to exchange the data with others. When data elements are well documented according to ISO/IEC 11179 and the documentation is managed in a Data Element Registry, finding and retrieving them from disparate databases as well as sending and receiving them via electronic communications are made easier.

The recognition and standardization of data elements used in communications through automated information processing systems is an ongoing and essential activity. The success of this activity and its application throughout the world is of vital importance if international communications among governments, businesses, and scientific communities are to be improved.

The primary data sharing and standardization problems addressed by the development of ISO/IEC 11179 include, but are not limited to the following:

- A lack of mechanisms for enabling global data acquisition and interchange, particularly across application areas;
- Unique global identifiers for standard data elements currently do not exist;
- Documentation of data element characteristics is inadequate to support fully automated sharing of data, including locating, retrieving, and exchanging the data;
- There is a lack of uniform guidance for identification, development, and description of data elements;
- Finding and retrieving a specific standard data element among thousands or millions is difficult or impossible;
- No universal means for organizing standard data elements exists;
- While data is sometimes standardized within an organization, there are few common data standards between organizations;
- Exchange of data among organizations results in a proliferation of customized data interchange representations;
- Data definitions and descriptions are not sufficiently precise to support reuse or multiple users of data;
- Current inventory structures for reducing logical data redundancies are inadequate;
- Global implementation of Electronic Data Interchange (EDI) is impeded by a lack of standard data elements; standard data elements are needed for the content of EDI messages.

To facilitate global electronic communications, the International Standards community has been working diligently to define an Open Systems Interconnection Environment (OSIE) within which diverse computer hardware and applications could share information. Standards have been proposed or defined for three (hardware, software, and

communications) of the four (hardware, software, communications, and data) basic components required for open information processing systems. ISO/IEC 11179 for data specification, the fourth basic component for open information systems, provides a mechanism for enabling data to be shared in the OSIE.

For systems to be truly open, data must be portable and shareable within and among these various application environments, which span localized and distributed networks. For data to be shareable, both the users and owners of data must have a common understanding of its meaning, representation, and identification. To understand the meaning of any data, the descriptions of the data must be available to the users from, for example, a Data Element Registry. Data must be adequately described and users must have a convenient way to obtain these descriptions. Data Element Registries provide a way to organize the content and representation of data elements so that data descriptions are consistently specified and can be easily located by data designers and users. Uniform specification of data facilitates data retrieval, data exchange, and consistent use of data throughout the Software Development Life Cycle. The units of information with normalized meanings and formats are known as "standardized data elements."

Purpose of ISO/IEC 11179

ISO/IEC 11179 describes the standardizing and registering of data elements to make data understandable and shareable. Data element standardization and registration as described in ISO/IEC 11179 allow the creation of a shared data environment in much less time and with much less effort than it takes for conventional data management methodologies.

The purpose of ISO/IEC 11179 is to give concrete guidance on the formulation and maintenance of discrete data element descriptions and semantic content (metadata) that shall be used to formulate data elements in a consistent, standard manner. It also provides guidance for establishing a data element registry.

Although motivated by the desire for the open exchange of data throughout the international communities by electronic information interchanges, ISO/IEC 11179:

- facilitates acquisition and registration of data;
- expedites access and use of data;
- simplifies data manipulation by intelligent software by enabling manipulation of data based on characteristics described by metadata;
- enables the development of a data representation metamodel for CASE tools and repositories; and
- facilitates electronic data interchange and data sharing.

ISO/IEC 11179 benefits the communication of data among information systems and people:

- within an organization;
- among different organizations; and
- crossing all levels of software and hardware, and geographic, organizational and political boundaries.

Metadata about data elements is stored in a data element registry. A data element registry supports data sharing with descriptions of data. Registration is the process of documenting metadata to support data shareability. Registration should be carried out at the data element level to promote and maximize semantic value. ISO/IEC 11179 enables the end user to interpret the intended meaning confidently, correctly, and unambiguously.

Users of ISO/IEC 11179

For users and managers of data, ISO/IEC 11179 specifies a basic set of data element characteristics necessary to share data. It places special emphasis on important data element characteristics such as identifiers, definitions, and classification categories. ISO/IEC 11179 describes a data element registry to assist users of shared data to have a common understanding of a data elements meaning, representation, and identification. If data values are received, the user can discover the exact meaning of the data received. If users wish to retrieve data values from a database, they can identify the type of data desired.

For systems analysts and data stewards, ISO/IEC 11179 provides a way to reuse a data element that meets a need, or to design a new data element if one does not already exist. Even before the user accesses data elements in a database, data stewards and systems analysts must have a way to identify and describe data logically so that they do not inadvertently introduce inconsistent values of data. If systems analysts are to create products that share data, they must first be aware whether or not a data element with the required characteristics already exists. If it does, they should use it. If the systems analysts choose to replicate the data element, they must represent data elements containing the same information in the same manner. If a data element with exactly the same characteristics does not already exist, a data steward needs to design the data element and make its description available to software developers. ISO/IEC 11179 aids in the development of precise descriptions of data elements. Data elements that have been formulated according to the principles in this multi-part International Standard enable interchangeability and retrieval regardless of the information processing system or telecommunication protocols employed.

For software developers, ISO/IEC 11179 provides means to assure data coherence. A registry can serve software developers by enabling the consistent use of data throughout the Software Development Life Cycle (SDLC). A registry will provide the mechanisms for managing data elements and for ensuring their traceability between SDLC phases.

For developers of a data dictionary, data element registry, CASE tool, and other data management software, ISO/IEC 11179 provides the basis for designing a metamodel necessary to enable the capture, storage, management, and exchange of the data element metadata.

1. Scope

ISO/IEC 11179 specifies basic aspects of data element composition, including metadata. It applies to formulation of data element representations and meaning as shared among people and machines; it does not apply to the physical representation of data as bits and bytes at the machine level.

This part of ISO/IEC 11179 provides the context for associating the individual parts and is the foundation for a conceptual understanding of data elements.

2. Normative References

The following standards contain provisions, which, through reference in the text, constitute provisions for this document. At the time of publication, the editions indicated were valid. All standards are subject to revision, and parties to agreements based on this document are encouraged to investigate the possibility of applying the most recent editions of standards indicated below. Members of IEC and ISO maintain registers of currently valid International Standards.

ISO Standards Handbook 10, Data Processing - Vocabulary, 1982

ISO 704:1987, *Principles and methods of terminology*

ISO 1087, *Terminology - Vocabulary*

ISO 2382-4:1987, *Information processing systems - Vocabulary part 4*

ISO/IEC 10241:1992, *International Terminology Standards - Preparation and layout*

FCD 11179-2, *Information technology - Specification and standardization of data elements - Part 2: Classification for data elements*

ISO/IEC 11179-3:1994, *Information technology - Specification and standardization of data elements - Part 3: Basic attributes of data elements*

ISO/IEC 11179-4:1995, *Information technology - Specification and standardization of data elements - Part 4: Rules and guidelines for the formulation of data definitions*

ISO/IEC 11179-5:1995, *Information technology - Specification and standardization of data elements - Part 5: Naming and identification principles for data elements*

ISO/IEC 11179-6:1997, *Information technology - Specification and standardization of data elements - Part 6: Registration of data elements*

3. Definitions

For the purposes of ISO/IEC 11179, the following terms are defined in the table below. An X under the column heading for a Part indicates that the term is defined in that Part and used in other clauses there. Each word that appears in **bold** in the definition of a term is a term defined elsewhere in this clause. Words that appear in regular type assume their commonly understood definitions. Some words (e.g. representation) are used both ways. There are instances where two or more terms appear next to each other in a definition, giving the appearance that a new term is undefined. There is no ambiguity in determining the actual terms in these cases.

Number	Term	Definition	Part Number					
			1	2	3	4	5	6
3.1	administered component:	A component for which administrative **attributes** are collected.		X				
3.2	administrative status:	A designation of the position in the processing life-cycle of a **registration authority** for handling **registration** requests.	X					X
3.3	attribute:	A characteristic of an **object** or **entity**.	X	X	X	X	X	X
3.4	attribute value:	A representation of an instance of an **attribute**.			X			
3.5	certified data element:	A **recorded data element** that has met the quality requirements specified in ISO/IEC 11179.	X					X
3.6	classification scheme:	An arrangement or division of **objects** into groups based on characteristics that the **objects** have in common, e.g., origin, composition, structure, application, function, etc.		X	X			X
3.7	classification scheme item:	A component of content in a **classification scheme**. This may be a node in a **taxonomy** or ontology, a term in a **thesaurus**, etc.		X				
3.8	classified component:	Any component of a **data element** that may be classified in one or more **classification schemes**. These components include the **object class**, **property**, **representation** class, **data element concept**, **value domain**, and **data element**.		X				

Number	Term	Definition	Part Number 1	2	3	4	5	6
3.9	comments:	Remarks on the **data element**.			X			X
3.10	concept:	A unit of thought constituted through abstraction on the basis of characteristics common to a set of **objects**. [ISO 1087]	X	X		X		
3.11	context:	A designation or description of the application environment or discipline in which a **name** is applied or from which it originates.			X		X	X
3.12	data:	A representation of facts, concepts, or instructions in a formalized manner, suitable for communication, interpretation, or processing by humans or by automatic means. [ISO 2382-4]	X	X		X		X
3.13	data dictionary:	A database used for **data** that refers to the use and structure of other **data**; that is, a database for the storage of **metadata** [ANSI X3.172-1990]. See also **data element dictionary**.	X	X		X		
3.14	data element:	A unit of **data** for which the **definition**, identification, **representation**, and permissible values are specified by means of a set of **attributes**.	X	X	X	X	X	X
3.15	data element concept:	A **concept** that can be represented in the form of a **data element**, described independently of any particular **representation**.	X	X	X		X	X
3.16	data element dictionary:	An information resource that lists and defines all relevant **data elements**. See also **register**.		X	X	X		
3.17	data element facet:	Any aspect of a **data element** that is subject to classification. This includes **object class**, **property**, **representation**, and **data element concept**.		X				

Number	Term	Definition	Part Number 1	2	3	4	5	6
3.18	data element name:	A single or multi-word designation used as the primary means of identification of **data elements** for humans.	X					
3.19	data element registry:	An information resource kept by a **registration authority** that describes the meaning and representational form of **data elements**, including registration **identifiers**, **definitions**, **names**, **value domains**, **metadata** and administrative **attributes**, etc. See also **register**.	X					
3.20	data element value:	A value out of a set of permissible values pertaining to a **data element**. See also **data value**.			X			X
3.21	data identifier (DI):	An **identifier** of a **data element** (a string of characters or other graphic symbols) assigned by a **registration authority**.					X	X
3.22	data item:	One occurrence of a **data element**.						X
3.23	data model:	A description of the organization of **data** in a manner that reflects an information structure.	X					
3.24	data steward:	A person or organization delegated the responsibility for managing a specific set of **data** resources.	X					
3.25	datatype:	The format used for the collection of letters, digits, and/or symbols, to depict values of a **data element**, determined by the operations that may be performed on the **data element**.	X					
3.26	datatype of data element values:	A set of distinct values for representing the **data element value**.			X			X
3.27	data value:	An element of a **value domain**.	X					

Number	Term	Definition	Part Number 1	2	3	4	5	6
3.28	definition:	A word or phrase expressing the essential nature of a person or thing or class of persons or things: an answer to the question "what is x?" or "what is an x?"; a statement of the meaning of a word or word group [Webster's Third New International Dictionary of the English Language Unabridged, 1986]. Statement that expresses the essential nature of a **data element** and permits its differentiation from all other **data elements**.	X		X	X	X	
3.29	domain:	The set of possible **data values** of an **attribute**. [ISO/IEC 2382]. See also **value domain**.	X	X		X		
3.30	entity:	Any concrete or abstract thing of interest, including associations among things. [ISO/IEC 2382]. Also see **object class**.	X	X				
3.31	enumerated domain:	A **value domain** that is specified by a list of all permissible values.	X					
3.32	form of representation:	Name or description of the form of **representation** for the **data element**. e.g. 'quantitative value, 'code', 'text', 'icon'. See also **representation term**.			X			X
3.33	identifier:	A language independent unique **identifier** of a **data element** within a **registration authority**. See also **data identifier**. An unambiguous name for an **object** within a given context.	X	X	X	X	X	X
3.34	information:	(In information processing): Knowledge concerning **objects**, such as facts, events, things, processes, or ideas, including **concepts**, that within a certain **context** has a particular meaning. [ISO/IEC 2382]	X					

Number	Term	Definition	Part Number 1	2	3	4	5	6
3.35	information interchange:	The process of sending and receiving **data** in such a manner that the **information** content or meaning assigned to the **data** is not altered during the transmission.	X					X
3.36	international registration data identifier (IRDI):	An internationally unique **identifier** for a **data element**.						X
3.37	keyword:	One or more significant words used for retrieval of **data elements**.		X	X			X
3.38	layout of representation:	The layout of characters in **data element values** expressed by a character string representation.			X			X
3.39	lexical:	Pertaining to words or the vocabulary of a language as distinguished from its grammar and construction.	X				X	
3.40	maximum size of data element values:	The maximum number of storage units (of the corresponding **datatype**) to represent the **data element value**.			X			X
3.41	metadata:	**Data** that defines and describes other **data**.	X					
3.42	minimum size of data element values:	The minimum number of storage units (of the corresponding **datatype**) to represent **the data element value**.			X			X
3.43	name:	The primary means of identification of **objects** and **concepts** for humans. A single or multi-word designation assigned to a **data element**.	X	X	X	X	X	X
3.44	object:	Any part of the conceivable or perceivable world. [ISO 1087].	X					

Number	Term	Definition	Part Number 1	2	3	4	5	6
3.45	object class:	A set of **objects**. A set of ideas, abstractions, or things in the real world that can be identified with explicit boundaries and meaning and whose **properties** and behavior follow the same rules.	X					
3.46	object class term:	A component of the **name** of a **data element** which represents the **object class** to which it belongs; e.g. "employee".	X				X	
3.47	permissible data element values:	The set of representations of permissible instances of the **data element**, according to the **representation form**, **layout**, **datatype**, **maximum size**, and **minimum size** specified in the corresponding **attributes**. The set can be specified by **name**, by reference to a source, by enumeration of the representation of the instances, or by rules for generating instances.			X			X
3.48	property:	A peculiarity common to all members of an **object class**.	X					
3.49	property term:	A component of the **data element name** which expresses a **property** of an **object class**. (A component of the **name** of a **data element** which expresses the category to which the **data element** belongs.)	X				X	
3.50	qualifier:	A term that helps define and render a **concept** unique.	X					
3.51	qualifier term:	A word or words which help define and differentiate a **name** within the database.					X	
3.52	recorded data element:	A submitted **data element** which contains all mandatory **attributes** and has been recorded but the contents may not meet the quality requirements specified in other parts of ISO/IEC 11179.	X					X

Number	Term	Definition	Part Number 1	2	3	4	5	6
3.53	register:	A set of files (paper, electronic, or a combination) containing the **assigned data elements** and the associated information. See also **data element registry**.						X
3.54	registration:	The assignment of an unambiguous **identifier** to a **data element** in a way that makes the **metadata** about those **data elements** available to interested parties.	X					X
3.55	registration applicant:	An organization, individual, etc, which requests the assignment of an **identifier** from a r**egistration authority**.	X					
3.56	registration authority (RA):	An organization authorized to register **data elements** or other **objects**.	X		X		X	X
3.57	registration authority identifier (RAI):	An **identifier** assigned to **a registration authority**.	X		X		X	X
3.58	registration status:	A designation of the position in the **registration** life-cycle of a **data element**.	X		X			X
3.59	related data reference:	A reference between a **data element** and any related **data**.			X			X
3.60	representation:	The combination of a **value domain**, **datatype**, and, if necessary, a unit of measure or a character set.	X					
3.61	representation category:	Type of symbol, character, or other designation used to represent a **data element**.			X			X
3.62	representation term:	A component of a **data element name** which describes the **form of representation** of the **data element**.	X				X	

Number	Term	Definition	Part Number 1	2	3	4	5	6
3.63	responsible organization:	The organization or unit within an organization that is responsible for the contents of the mandatory **attributes** by which the **data element** is specified.			X			X
3.64	semantics:	The branch of linguistic science which deals with the meaning of words (Webster).	X				X	
3.65	separator:	A symbol or space enclosing or separating components within a **name**; a delimiter.					X	
3.66	standardized data element:	A **certified data element** within the **data element registry** that is preferred for use.	X					X
3.67	structure set:	A method of placing **objects** in **context**, revealing relationships to other **objects**. Examples include entity-relationship models, **taxonomies**, and ontologies.					X	
3.68	submitting organization (SO):	The organization or unit within an organization that has submitted the **data element** for addition, change, cancellation, or withdrawal in the **data element registry**.			X			X
3.69	synonymous name:	Single or multi-word designation that differs from the given **name**, but represents the same **data element concept**.			X			X
3.70	syntax:	The structure of expressions in a language, and the rules governing the structure of a language. The relationships among characters or groups of characters, independent of their meanings or the manner of their interpretation and use.	X				X	
3.71	taxonomy:	Classification according to presumed natural relationships among types and their subtypes.	X	X				

Number	Term	Definition	Part Number					
			1	2	3	4	5	6
3.72	term:	Designation of a defined concept in a special language by a linguistic expression. [ISO/IEC 1087]	X					
3.73	thesaurus:	A controlled vocabulary arranged in a given order in which relationships among terms are displayed and identified.	X				X	
3.74	type of relationship:	An expression that characterizes the relationship between the **data element** and related **data**.			X			X
3.75	value domain:	A set of permissible values.	X					
3.76	version:	The identification of an issue of a **data element** specification in a series of **evolving data element** specifications within a **registration authority**.			X			X
3.77	version identifier (VI):	An **identifier** assigned to a **version** under which a **data element registration** is submitted or updated. NOTE: Part 5, same as **version**.			X		X	X

4. Abbreviations

CASE -- Computer-Aided Software Engineering

EDI -- Electronic Data Interchange

ERD -- Entity-relationship Diagram

IEC -- International Electrotechnical Commission

ISO -- International Organization for Standardization

JTC1 -- Joint Technical Committee 1

OSIE -- Open Systems Interconnection Environment

RA -- Registration Authority

RDBMS -- Relational Database Management System

SC32 -- ISO/IEC JTC1/Sub-committee 32

SDLC -- Software Development Life Cycle

5. Methodology Independence

It is recognized that different methodologies are used to derive application-oriented data elements. Data mo deling, alone, has numerous methodological approaches (e.g., Information Engineering and object-oriented) for identifying and forming data elements. This data specification standard, including its six parts, is independent of any data element derivation methodology or technique. Since the standard applies to all data elements, it can be used equally well in any approach to data element creation.

6. Fundamental Concepts of Data Elements

For the purposes of ISO/IEC 11179, a data element is composed of three parts as follows:

- the **object class** is a set of ideas, abstractions, or things in the real world that can be identified with explicit boundaries and meaning and whose properties and behavior follow the same rules;

- the **property** is a peculiarity common to all members of an object class; and

- the **representation** describes how the data are represented, i.e. the combination of a value domain, datatype, and, if necessary, a unit of measure or a character set.

Object classes are the things about which we wish to collect and store data. Examples of object classes are cars, persons, households, employees, orders, etc. However, it is important to distinguish the actual object class from its name. Ideas simply expressed in one natural language (English), may be more difficult in another (Chinese), and vice-versa. For example, "women between the ages of 15 and 45 who have had at least one live birth in the last 12 months" is a valid object class not easily named in English. Nevertheless, object classes can be formed by combining two or more other object classes. This example combines the notions of "people between the ages of 15 and 45" with "women who have had at least one live birth in the last 12 months".

Properties are what humans use to distinguish or describe objects. Examples of properties are color, model, sex, age,

income, address, price, etc. Again, properties may need to be described using multiple words, depending on the natural language in use.

The most important aspect of the representation part of a data element is the value domain. A **value domain** is a set of permissible (or valid) values for a data element. For example, the data element representing annual household income[1] may have the set of non-negative integers (with units of dollars) as a set of valid values. This is an example of a **non-enumerated domain**. Alternatively, the valid values may be a pre-specified list of categories with some identifier for each category, such as:

1	$0	- $15,000
2	$15,001	- $30,000
3	$30,001	- $60,000
4	$60,001	- +

This value domain is an example of an **enumerated domain**. In both cases, the same object class and property combination - the annual income for a household - is being measured.

The combination of an object class and a property is a **data element concept** (DEC). A DEC is a concept that can be represented in the form of a data element, described independently of any particular representation. In the examples above, annual household income actually names a DEC, which has two possible representations associated with it. Therefore, a data element can also be seen to be composed of two parts: a data element concept and a representation.

Figure 6-1 illustrates the ideas conveyed in this section.

[1] Part 5 of ISO/IEC 11179 specifies the inclusion of a representation term in the name of a data element, so annual household income is not a complete data element name. The omission is intentional.

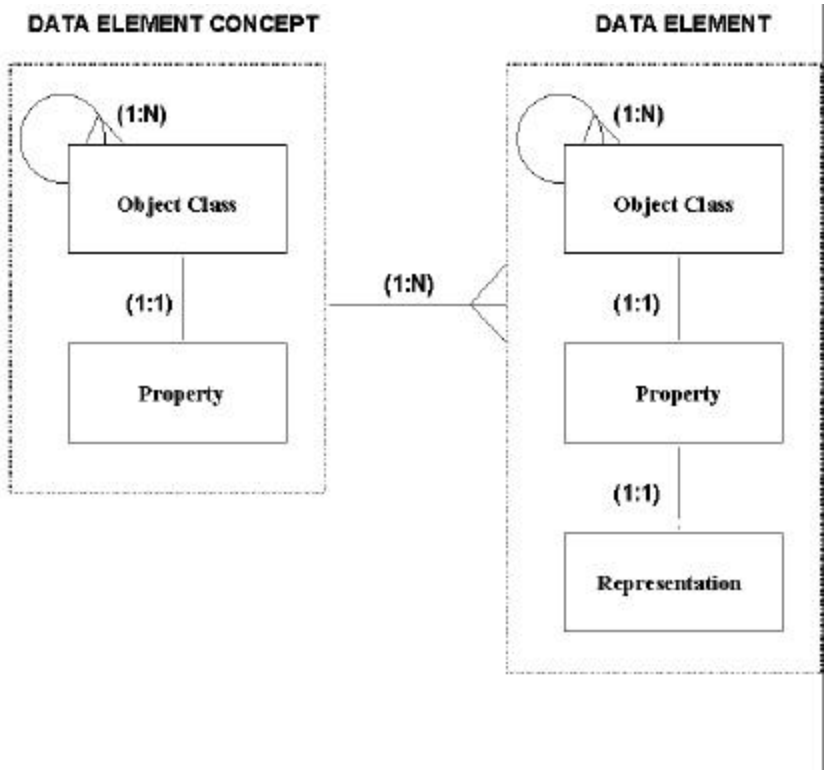


Figure 6-1: Data Element Structure

7. Relationship of Data Elements to Other Data Concepts

Figure 7-1 provides a simplified representation of levels of data to illustrate those levels in which data elements lie. Data elements appear in databases, files, and transactions sets. Data elements are the fundamental units of data an organization manages, therefore they must be part of the design of databases and files within the organization and all transaction sets the organization builds to communicate data to other organizations.

Within the organization, databases or files are composed of records, segments, tuples, etc. which are composed of data elements. The data elements themselves contain various kinds of data that include characters, images, sound, etc.

When the organization needs to transfer data to another organization, data elements are the fundamental units that make up the transactions sets. Transactions occur primarily between databases or files, but the structure (i.e. the records or tuples) of the files and databases don't have to be the same across organizations. So, the common unit for transferring information (data plus understanding) is the data element.

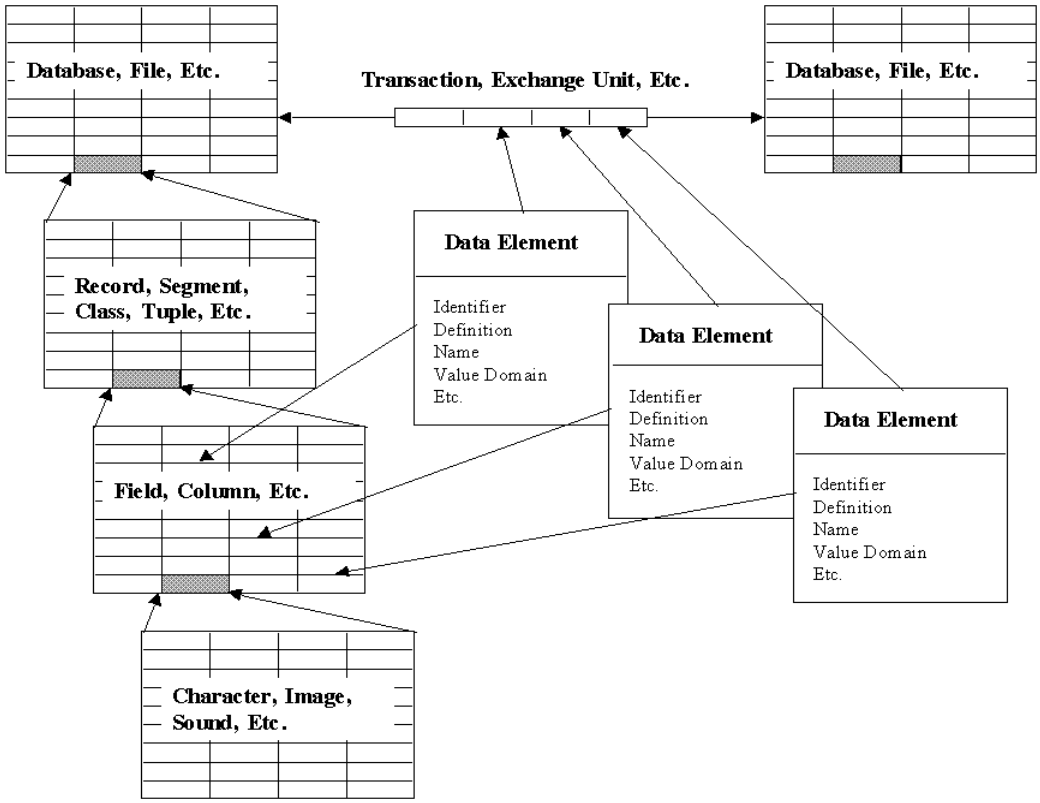


Figure 7-1: Data Elements in Levels of Data

8. Overview of ISO/IEC 11179, Parts 1-6

8.1. Introduction of Parts

This clause introduces each part of the multi-part standard ISO/IEC 11179. It summarizes the main points and discusses the importance of each.

8.1.1 Part 1

ISO/IEC 11179-1, Framework for the Specification and Standardization of Data Elements, introduces and discusses fundamental ideas of data elements essential to the understanding of this set of standards and provides the context for associating the individual parts of ISO/IEC 11179.

8.1.2 Part 2

Classification for Data Elements, ISO/IEC 11179-2, provides procedures and techniques for associating data element concepts and data elements with classification schemes for object classes, properties and representations.

These procedures and techniques shall assist Registration Authorities in applying classification schemes that enable

them to perform activities such as:

- analyze object classes, data element concepts, and data elements;
- make comparisons within the following categories: object classes, data element concepts, and data elements;
- reduce the variety of data element concepts and data elements;
- define and identify data element concepts and data elements unambiguously;
- assist in the analysis of data elements for the purpose of assigning registration status;
- retrieve data element concepts and data elements from a data element registry;
- recognize relationships among data element concepts and data elements.

ISO/IEC 11179-2 develops a set of principles, methods, and procedures for specifying what is needed (at a minimum) in a taxonomy/ontology for description of object class, property, representation, data element concepts, and data elements (herein called "information elements"). This includes the need for names, non-intelligent identifiers, definitions, and other things that can be captured as attributes. ISO/IEC 11179-2 shows how to go from a taxonomy to attributes and how to use the attributes specified in ISO/IEC 11179-3 as well as other attributes that are needed. Taxonomies are provided in an informative annex. A suggested set of qualifiers that could be applied to the property information element and a set of qualifiers for the representation information element are provided. The qualifier sets are accompanied with an explanation of how to address synonym and homonym problems. ISO/IEC 11179-2 also shows how information elements relate to taxonomies and how the pertinent taxonomic information can be used to describe the above information elements.

8.1.3 Part 3

Basic Attributes of Data Elements, ISO/IEC 11179-3, specifies attributes of data elements. It is limited to a set of basic attributes for data elements, independent of their usage in application systems, databases, data interchange messages, etc.

The increased use of data processing and electronic data interchange heavily relies on accurate, reliable, controllable, and verifiable data recorded in databases. One of the prerequisites for a correct and proper use and interpretation of data is that both users and owners of data have a common understanding of the meaning and representation of the data elements. To guarantee a shared view of data elements, a number of basic attributes have to be defined.

The basic attributes specified are applicable for the following main activities:

a) definition and specification of the contents of data element dictionaries;

b) design and specification of application-oriented data models, databases, and messages for data interchange;

c) actual use of data in communications and information processing systems;

d) interchanging or referencing among various collections of data elements.

Basic means that they are essential in specifying a data element completely enough to ensure that it will be applicable for a variety of functions, such as:

- design of information processing systems;

- retrieval of data from databases;
- design of EDI-messages for data interchange;
- maintenance of data element dictionaries;
- data management;
- dictionary design;
- dictionary control;
- use of information processing systems.

Basic also implies that they are independent of any:

- application environment;
- function of a data element (e.g., qualifier, indicator);
- level of abstraction of the meaning (e.g., a representation of a generic concept like 'name of a person' or a representation of a specific concept like 'name of the driver of a truck');
- grouping of data elements;
- method for designing information processing systems or data interchange messages;
- data element registry system.

Basic does not imply that all attributes specified in ISO/IEC 11179-3 are required in all cases. Distinction is made between those basic attributes that are:

- mandatory - always required[2];
- conditional: required to be present under certain specified conditions;
- optional - allowed but not required.

8.1.4 Part 4

Rules and Guidelines for the Formulation of Data Definitions, ISO/IEC 11179-4, provides guidance on how to develop unambiguous data element definitions. A number of specific rules and guidelines are presented in ISO/IEC 11179-4 that specify exactly how a data element definition should be formed. A precise, well-formed definition is one of the most critical requirements for shared understanding of a data element; well-formed definitions are imperative for the exchange of information. Only if every user has a common and exact understanding of the data element can it be exchanged trouble-free.

8.1.5 Part 5

Naming and Identification Principles for Data Elements, ISO/IEC 11179-5, provides guidance for the identification of

[2] Required to record data elements. ISO/IEC 11179-3 does not address recording the subsets of attributes required to record data element concepts.

data elements. Identification is a broad term for designating, or identifying, a particular data element. Identification can be accomplished in various ways, depending upon the use of the identifier. Identification includes the assignment of numerical identifiers that have no inherent meanings to humans; icons (graphic symbols to which meaning has been assigned); and names with embedded meaning, usually for human understanding, that are associated with the data element's definition and value domain.

Names are semantic, natural language labels given to data elements, and variations of these labels serve different functions. Some names are for human usage and comprehension; some names are for use in a particular physical system environment. Names are often user established and vary from one user to the other. The principles in ISO/IEC 11179-5 describe the various functions of names and how names are used. One and only one identifier is required for each data element within a Registration Authority. The identifier does not change as long as the meaning and type of representation of the data element remains unchanged. Identifiers are unique only within a Registration Authority. ISO/IEC 11179 does not specify the format or the content of identifiers.

Each data element may be given many names - each with meaning within a particular context. Structured names may be registered for data management, a preferred name may be specified by the enterprise, many common names may be familiar to various user groups, shortened names may be registered for software applications, etc. A naming convention is enforced within each context. This convention may be different between contexts. Along with each name for a data element, the description of the context must be provided.

One type of structured name may be derived from the classification scheme described in ISO/IEC 11179-2. The components of this name come from the three classification taxonomies: Object Class, Property, and Representation. By assembling the name of each taxon along with some descriptive qualifiers, a data element name can be built. This name can be valuable for assisting the user to find or place a data element in a taxonomy intuitively. However, it is not invariant and is not dependable as a unique identifier.

ISO/IEC 11179-5 addresses identifying and naming data elements. Although the same principles should apply, it does not specifically address names or identifiers for representations, properties, or object classes.

8.1.6 Part 6

Registration of Data Elements, ISO/IEC 11179-6, provides instruction on how a registration applicant may register a data element with a central Registration Authority and the allocation of unique identifiers for each data element. Maintenance of data elements already registered is also specified in this document.

The uniqueness of a registered data element is determined by the combination of the Registration Authority Identifier, the unique identifier assigned to a data element within a Registration Authority, and the version. They are also included in widely available Registries of data elements. Each registry is maintained by a Registration Authority to which data elements logically and functionally belong. For example, data elements related to chemical matter would likely be registered under a Chemical Manufacturer Registration Authority. The Registries should be indexed and constructed so that those designing applications or messages, such as EDI, can ascertain easily whether a suitable data element already exists. Where it is established that a new data element is essential, the procedure should encourage its derivation from an existing entry with appropriate modifications, thus avoiding unnecessary variations in the way similar data elements are constructed. Registration will also allow two or more data elements serving an identical function to be identified, and more importantly, it will identify situations where similar or identical names are in use for data elements that are significantly different in one or more respects.

Registration is more complex than a binary status simply indicating whether a data element is either registered or not. Although it is tempting to insist that only "good" data may be registered, that is not practical. Therefore, improvement in the quality of registered data is divided into three levels (called registration status): recorded data element, certified data element, and standardized data element. In addition, there are status levels for administration between each of these quality levels. Collectively, these status levels are called administrative status. They indicate the point in the registration life cycle currently attained by a registered data representation.

8.2 Basic Principles for Applying ISO/IEC 11179, Parts 1-6.

Each Part of ISO/IEC 11179 assists in a different aspect of data element formulation, and each Part shall be used in conjunction with the other Parts. ISO/IEC 11179-1, *Framework*, establishes the relationships among the Parts and gives guidance on their usage as a whole. ISO/IEC 11179-3, *Basic Attributes of Data Elements*, specifies a set of mandatory metadata items a registration applicant shall provide for each data element. In addition, a list of potential additional items for use as needed is provided. Detailed characteristics of each basic attribute are given. Because of their importance to a data element, two of the mandatory attributes, definition and identification, are given special and extensive treatment in separate documents: ISO/IEC 11179-4, *Rules and Guidelines for the Formulation of Data Definitions* and ISO/IEC 11179-5, *Naming and Identification Principles for Data Elements*.

The document on definitions shall be followed when constructing data element definitions. A data element's identification shall follow principles set forth in ISO/IEC 11179-5. ISO/IEC 11179-2, *Classification for Data Elements*, specifies a set of attributes for use in development of classification schemes for data elements and their components. On completion of a data element's formulation, it can be registered with a Registration Authority where it shall be maintained in a Data Element Registry. ISO/IEC 11179-6, *Registration of Data Elements*, provides guidance on these procedures.

Annex A (informative)

Fundamental Concepts of Data Representation and Management

This annex introduces the fundamental ideas upon which ISO/IEC 11179, Parts 2-6, have been based.

A.1 Data Elements

A data element is a special member of the broad notion called data. In general, data is a **representation** of facts, ideas, or instructions. Data is collected, organized, recorded, processed, and stored in a retrievable form. Data must also be suitable for communication, interpretation, or processing by human or automated means.

There are many constructs used to organize data. There are data composites, entities, files, object classes, objects, records, relations, relationships, rows, segments, subject areas, tables, and tuples. None of these are analogous to data elements, but may include or be supported by some database implementation or logical modeling equivalent of data elements.

Bytes and bits are also components of data. Although they may be used to record data elements in an electronic medium, they do not correspond to data elements. In a database, a data element may be implemented as a field or column. In Chen's ER data model, it is an attribute (see Figure A-5). A data element then is a single unit of data that in a certain context is considered indivisible. It is a unit of data representing a single fact about a type of object (object class) in the natural world. (For example, a one character code with allowed values of "M" or "S" representing the marital status attribute of an "employee" object class.) It cannot be decomposed into more fundamental segments of information that have useful meanings within the scope of its application. Data elements are thus defined as relevant to the user within the user's universe of discourse. Data elements are electronic or written representations of the properties of natural-world object classes.

A.1.1 Representation

A property may be incarnated by some symbol set that can be interpreted by humans. A single property can be represented by several alternate data elements or even groups of data elements (usually called a data composite or occasionally a data aggregate or chained data elements). A property can thus be symbolized by either a data composite made up of a group of data elements, or a single data element.

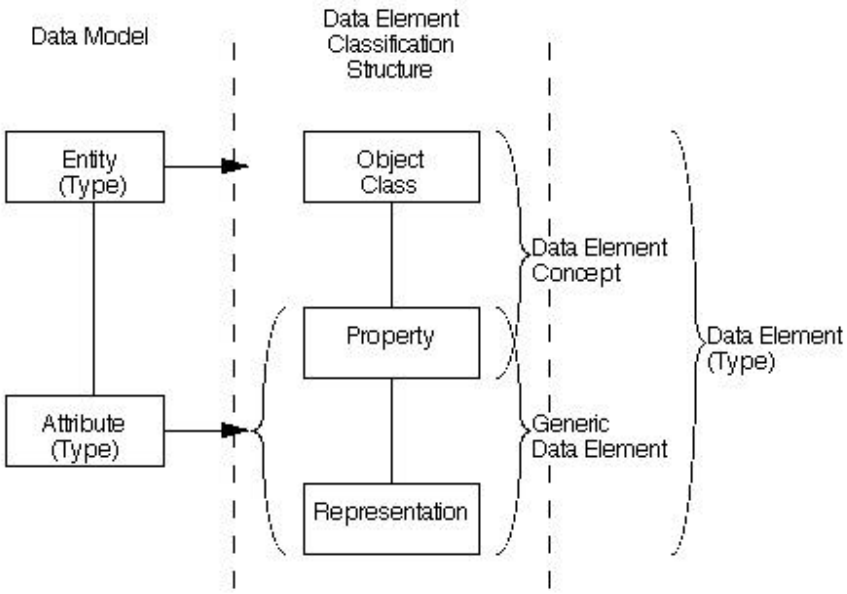
A.1.2 Common Usage

All data practitioners and theorists can relate to the idea of the data element. The data element is the common link throughout the Software Development Life Cycle (SDLC), no matter what methodology or technology is used. In one of the earliest stages of the SDLC, they are considered attributes of entities (or object classes). Near the end of the SDLC, specific data values are assigned to instances of them as fields or columns. In any phase of the SDLC, data elements are recognizable to software producers and users.

The data element is the smallest unit of data that is shared and held in common. Some of the information system components among which data elements are shared are:

- Enterprise Information Model;
- Data Models;
- Data Flow Diagrams;

Figure A-1: Data Element Structure



- Database Designs (schemas, files, tables);
- Interface Specifications;
- Computer Programs.

A.1.2.1 Data Elements in Data Models

The term "data element" was in common use long before the arrival of data modeling. The units of data used to represent the information needed to conduct the business of an enterprise have traditionally been called data elements. Data modeling evolved to provide an efficient way to capture the semantics of these data representations.

Figure A-1 depicts how the data element construct and terms used in ISO/IEC 11179 may relate to some more traditional data-modeling terminology.

In a data model, an attribute is a characteristic of an entity (entity type, object class, etc.) that the enterprise chooses to record as data. For each entity, there are often many attributes of interest to the enterprise.

Data models and object models (in the object-oriented paradigm) identify application-context things of interest (entities or objects). Attributes provide information required by the application about these entities and objects. Data elements for use throughout the automated information system environment are representations of these entities or objects and their attributes. Data elements names created from the data model are typically composites of the entity name and the entity attributes names (Figure A-2).

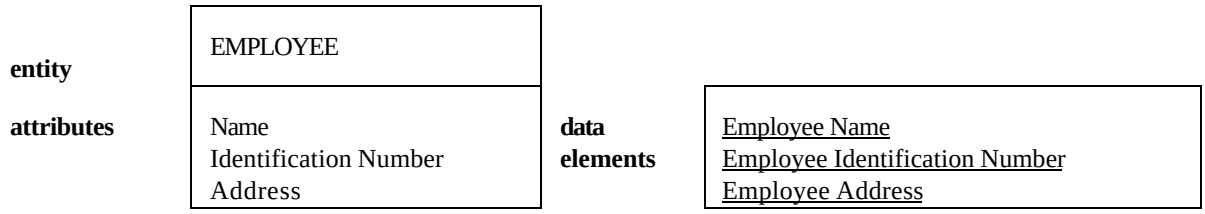


Figure A-2: Data Elements in a Data Model

Names of data elements in the object-oriented paradigm are generally created in the same manner. In the object model, class or object names are used in combination with the class or object attributes to produce data element names (Figure A-3). Object models differ from data models in that they may contain additional information about the object or class, such as behavior or operations.

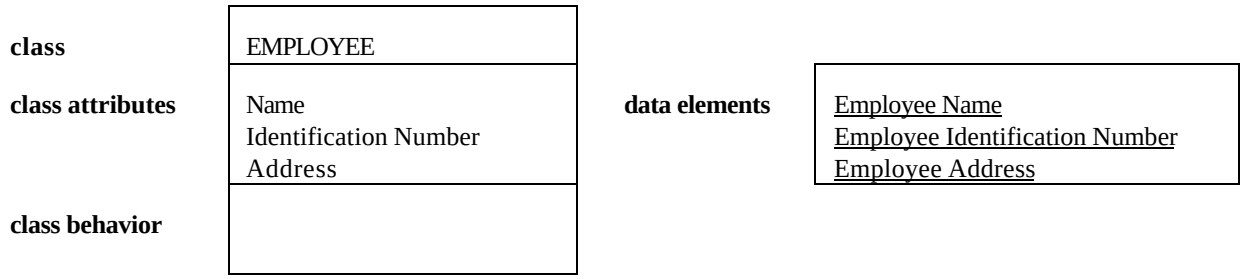


Figure A-3: Data Elements in an Object Model

In a data model or object model, an attribute is thus usually equivalent to a data element (see Figure A-5). It is the representation of a single property of a class of objects in the natural world. However, another school of thought separates the idea of the property of the object class from the representation of the property. Since data elements include representations, it is reasonable and enlightening to isolate the underlying idea that can be grasped by the human mind without relying upon how it is represented in a database, on a screen, on paper, etc. This is consistent with some data-modeling techniques that recognize an attribute as being **associated** with a value domain rather than an attribute **having** a value domain.

In some cases, such as units of measure, there is clearly a difference between an attribute and a data element. For example, "date" is a single point on a time continuum measured to the nearest day. As such, it can be considered a single fact and thus be represented as a single attribute in a data model. However, there are several ways to represent a date. In the U.S., the most common way is: name of month, day of month count, and year count. These are clearly three data elements. However, the same date fact can be represented by a single data element by using the number of days elapsed since some predetermined past date (e.g., Julian Day used by astronomers). Many other examples of single facts that can be represented in more than one way can be demonstrated using units of measure, in one case represented in the English system or the same fact using the metric system.

The conceptual equivalent of a data element has been called: attribute, property, data element concept, logical data element, and business fact. Herein it is referred to as a property to distinguish it from what is often known as an attribute (which usually includes its representational form). **A property, then, is a peculiarity common to all members of an object class.**

The property can be explained with an example. Suppose a tree is a natural-world object in which we are interested. However, we are concerned with any tree, not just a particular tree. The characteristic of trees in which we are

interested is their height. **Tree height** is an object class plus a property (data element concept), but not yet a data element, since the appropriate representational form has not yet been specified. We can then represent the height of a tree by choosing from among the many ways of measuring height.

The term property class might be preferred over property to name that aspect of a data element. A class of objects (object class), e.g. people, doesn't have height; each individual object, i.e. person, has a height. So for the object class called people, height is a property class of that object class. However, the term property conforms to common usage, and is used in this standard.

Sometimes data elements are derived from several constituent parts, where each of the parts are represented as data elements. These derivations can be of many forms. An example is concatenation for the formation of a telephone number from its constituent parts. In the U.S., telephone numbers are uniquely described with ten digits, and these numbers can easily be represented by a data element. However, the telephone companies (and others) need the telephone number separated into area code, exchange code, and line number, making three data elements. Concatenating the area code, exchange code, and line number (in the right way) allows the formation of a data element representing the full telephone number.

Another example of a derivation is an algebraic manipulation. The calculation of a mean (or average) requires the observations and the number of observations. Calculating the average age of persons in a set of people therefore requires a data element representing the ages of persons in the set and a data element representing the number of persons in the set. Then, the obvious arithmetic is carried out through a formula specified by the derivation.

These examples are fairly simple, but it is not hard to imagine much more complicated types of derivations that need to be captured in data elements. The derivation associated with a data element constitutes a powerful way of associating existing data elements with new ones that are needed.

A.1.2.2 Data Elements in Tables

Data elements in relational databases appear as field labels in tables. Figure A-4 provides an example of data elements in database tables.

record	EMPLOYEE			
attribute	Number	Last Name	Birth Date	Salary Amount
data values	1	Rood	47/3/4	483.00
	2	Herden	48/6/3	501.00
	3	Albright	51/7/9	490.00

Figure A-4: Data Elements in a Database Table

A.1.2.3 Data Elements in Data Management Tools

The following table (Figure A-5) identifies terms in data management tools that are usually associated with data elements (those in bold caps).

traditional	file	record	**FIELD**	field value
relational	relation	tuple	**ATTRIBUTE**	element
object-oriented	__	class	**ATTRIBUTE**	instance
RDBMS*	table	row	**COLUMN**	data value
ERD**	__	entity	**ATTRIBUTE**	__
Alternate ERD Terminology	__	entity type	**ATTRIBUTE TYPE**	__

*relational database systems
**Chen style entity-relationship diagrams

Figure A-5: Data Management Terms as Data Elements

A.2 Classification

A.2.1 Thesaurus

A thesaurus is a tool that associates related terms, and thesaurus terms assist in locating an existing data element. Finding a variety of synonyms, near-synonyms, and homographs for name components makes a thesaurus a valuable tool. It can provide semantic linking between preferred name terms and other terms. In addition to guidance for use of homographs (words with the same spelling representing different concepts), a thesaurus can direct the user through choices involving equivalence and hierarchical and associative relationships.

A thesaurus for the components in a standard name may be developed and distributed to interested parties by the registrar; in addition, development of subject area thesauri shall be encouraged.

A.2.2 Taxonomies and Ontologies

Classification (e.g., genus and species) could be done in at least two ways. The first, and simplest, is to standardize a method by which a classification is created each time a new data element is registered. The ideal, but most elaborate way is to start with a complete taxonomy of all possible data elements and then place each newly registered data element in its predefined position. The most practical approach is to start with a basic and relatively simple taxonomy and allow it to evolve, using rigorous rules, as data elements are registered.

The most common approach to creating a taxonomy is through lexicography. The lexical procedure focuses on the words or vocabulary of a language rather than directly upon the semantics. A lexicographer's approach serves the needs of a data element taxonomy if one accepts the premise that humans cannot think about a concept unless they have words to describe it. That is, humans have invented words for any concept about which they think.

Each data element registry should probably have a specific taxonomy that best meets the needs of its clients. Thus different registries may choose to use different taxonomies.

In a taxonomy, each node is a sub-type of one or more super-types. Not only does the node inherit the meaning of the super-type, the meaning is at the same time constrained by the super-type. No matter how well the definition of a data element is written, the classification greatly assists in exactly delineating the meanings of the node used in a classification. Another major benefit a taxonomy provides is for searching for a specific data element. The taxonomy

supports navigation through massive amounts of data element descriptions.

There is a basic classification structure that lends itself nicely to data element registration, analysis, and use. It is based on the premise that a data element is the **representation** of a **property** of an **object class** that exist in the natural world. The classification of a data element is the union of these three classifications: 1) the object class, 2) the property, and 3) the representational form. By searching the labels for each of these concepts, the desired data element can be located.

A data element classification scheme can best be described with a scenario. However, there is more than one scenario for using a classification scheme for data elements. It largely depends upon whether the metadata user is an end-user of data, or someone involved in Information Resource Management. If the latter, it depends upon the stage of the Software Development Life Cycle in which they are involved. One possible scenario is its use by a data modeler when defining requirements for a database structure to support a specific application:

1) The data modeler discovers the need for this application to process a specific item of data. In this case, it is the measurement of the height of horses.

2) The data modeler recognizes that the "property" could be called "height." The data element registry is consulted using the PROPERTY TAXONOMY for navigation through the registry. In the area of the taxonomy dealing with size, a taxon called "Height" is found. The definition of this taxon confirms that this describes the intended property.

3) The modeler recognizes that the object class could be called "horse." The OBJECT CLASS TAXONOMY in the registry is consulted and a taxon named "Horse" is found somewhat below "Animal." The definition of the taxon confirms that this is the one desired. HORSE will become the object class in the data model for the application.

4) The data element registry reveals that height is a property that has been previously identified as being associated with the horse object class. The registry says that this property for a horse is the distance from the surface upon which a horse is standing to its withers (i.e., shoulder).

5) The REPRESENTATION TAXONOMY is used to search the registry for a measurement. A standardized data element is located in the registry that is a metric measurement of horse height in METERS. However, the data modeler knows that this data element will not serve. The end-user of this application insists that horse heights be measured and recorded in the traditional system of HANDS. Searching the registry using the classification scheme discloses that no such data element has been registered.

6) The data modeler describes the representational form of a new data element that measures any horse's height in hands as the unit of measure. This data element is fully described and is submitted for registration. Only its representation (i.e., mostly value domain) needs to be newly described since the horse object class and its dependent height property has already been registered.

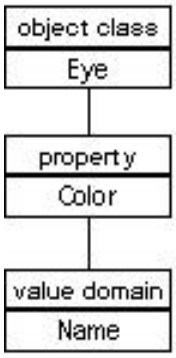
7) The HEIGHT HANDS MEASURE will become an attribute of horse in the application data model.

8) In the future, anyone needing this data element will find its description in the registry by using the three taxonomies of the classification scheme. The Horse taxon in the Object Class Taxonomy linked to the Height taxon in the Property Taxonomy linked to the Hands Measurement taxon in the Representation Taxonomy will lead any future user directly to the description of this data element.

The classification scheme for data elements is in detail in ISO/IEC 11179, Part 2, *Classification Principles for Data Elements.*

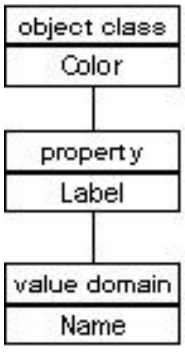
In practice, the difference between a property and an object class is often not absolute. The distinction is associated with the universe of discourse under consideration. Examples of physical properties are the best way to demonstrate this. For example, Eye Color Name can be considered a data element concept with Eye being the object class and Color the property. One of the possible data elements stemming from this data element concept could be Eye Color Name.

EXAMPLE 1



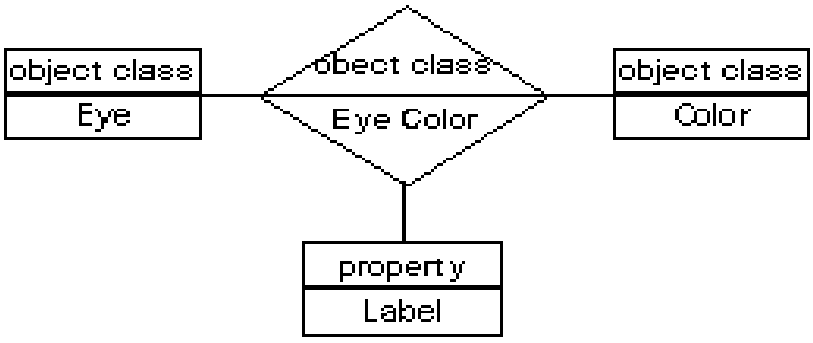
However, if one was dealing with labeling the various observable colors, Color would be the object class and the property could be called Label. One of the associated data elements could be called Color Name. Here, Color is an object class where in the first example it was a property. Neither way can be considered incorrect.

EXAMPLE 2



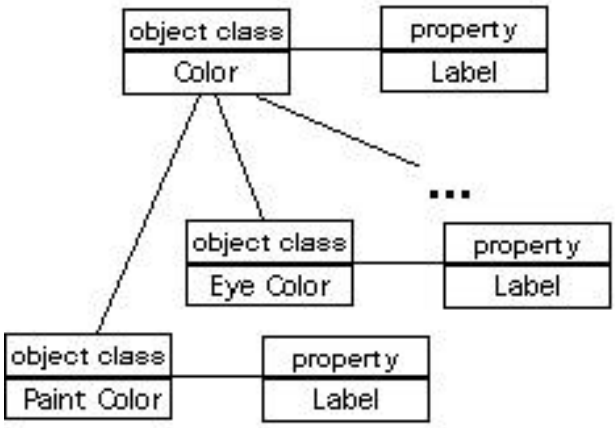
This second construct could be used for Eye Color by recognizing Eye Color as an object class that is a relationship between the object class Eye and the object class Color.

EXAMPLE 3



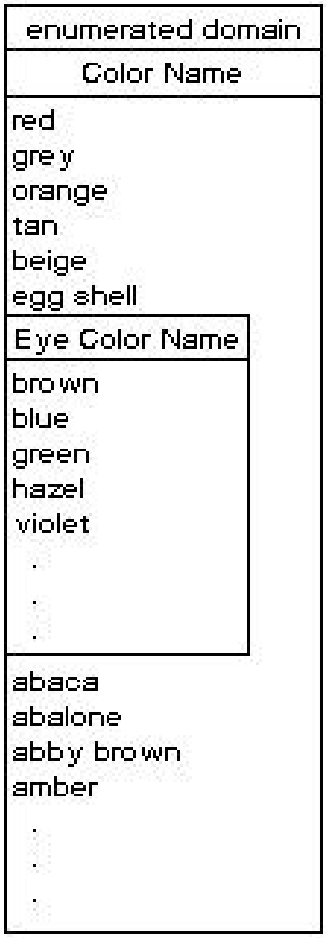
Yet another way to depict Eye Color is as a subtype. Eye Color could be an object class that is a sub-class of the object class Color.

EXAMPLE 4



The practical results of examples 3 and 4 are virtually identical. In the case of both examples, the value domain of Eye Color Name would be a subset of the value domain of Color Name.

EXAMPLE 5



Another way of considering this is that the value domain of Color Name is a super-set of all value domains of 1) in the case of example 4, the sub-classes of Color, or 2) in the case of example 3, all the object classes that are relationships to Color. ISO/IEC 11179 allows any of these data element construct variations to be used.

A.3 Data Element Metadata Attributes

A.3.1 Identification

Unique numerical identifiers are assigned to each registered data element as well as the property and object class represented by the data element. In some cases, unique numerical identifiers are also given **data values** of a data element. These unique identifiers are registered in a data element registry. They enable direct translation of the data elements across natural languages and application systems. Although data elements will have many different translations into different languages or multiple versions in a single language, the unique identifier will remain

constant and be the bridge for translating from one version to another.

Each **data element** shall receive an identifier which identifies the data element uniquely throughout any computing space, e.g., application program, local computing system, distributed computing system, organization, industry, and globally for all nations. Identifiers contain no intelligence. They can therefore be assigned to data elements by the registrar and remain permanently attached to them. Identifiers for exchange of data are assigned and maintained by a Registration Authority.
In addition to an identifier, a data element may be assigned any number of alternate names or icons bounded by disciplines or subject areas. One data element name may be a preferred name within a Registration Authority. Like the identifier, the preferred name maintains a one-to-one relationship with the data element. Names are usually the primary means by which human users of data identify and interact with data elements.

Data elements are registered by different Registration Authorities. Each of these may prefer to retain the identifiers, names, icons or other means of identification with which the users of the data are familiar. For alternate identifiers, the identifier coupled with the context in which this identifier has meaning is documented.

Each **data value** in some value domain may also be assigned an identifier. This is particularly useful for enumerated domains, as it can facilitate international language translation for associated names of data instances. For example, the data element "eye color name" may have the enumerated domain of "gray, blue, green, and brown," where each data value would have a unique identifier of its own (e.g., gray = 1357; blue = 2468).

A.3.2 Definition

A definition is a natural language statement of the data element's meaning; its predicate. The definition of a data element is an extremely critical aspect of data element development. To be shareable, a data element must have a well-formed, unambiguous, precise, and commonly understood definition. An extensive discussion of data element definitions is contained in ISO/IEC 11179, Part 4, with precise rules and guidelines to their formulation.

A.3.3 Representation

Data modelers often call the representational form of an attribute its "value domain" or simply "domain." For the purposes of ISO/IEC 11179, we refer to the representational part of a data element as **representation**.

Data elements take on values in business operations that can be shared among functions in the exchange of information. A data element always has a set of allowed data values. The set of allowed values is called a **value domain**.

A data element is never represented by a single data value since it is a type (i.e., the complete set of data values) and not a single instance. For example, Employee Identifier is a data element that has a value domain described by a complete list of data values allowable within a particular enterprise. The data values here would simply be a list of all instances of employee identifiers. An instance of a data element has a single data value and is called a "data element instance."

A property has a definition and belongs to an object class. Conversely, a representation has no definition but has a format category, permissible data values, maximum character count, and, if a measurement, a unit of measure. For example, the "name of month" data element has the following representation:

- Format category = alpha
- Permissible data values = January, February, March, April, May, June, July, August, September, October, November, December
- Maximum characters = 9

- Unit of measure = (not applicable since this is not a measurement)

. **A.3.3.1 Enumerated Domains**

Although the structure of data element metadata can be viewed from the perspective of an object class or a property, it can also be viewed from the perspective of a representation. This is easiest to illustrate with an *enumerated domain* example. An enumerated domain is a value domain that is specified by a list of all allowed values. For example, humans have labeled countries with names so that they can conveniently refer to them. The set containing the names of all the current countries in the world is an example of an enumerated domain.

Whereas the property and an associated value domain (of a representation) can be considered to identify the set of potential valid data values for a data element, the object class usually restricts or limits that extensive value domain to a specific subset of data values that apply specifically to a particular data element. Figure **A**-6 illustrates the restriction of value domains by different object classes

.In the data element example in Figure A-6, Name of Person Employed by Corporation X:

- "Name" has the value domain of all conceivable names,
- "of Person" refers to a subset of all possible names, and
- "employed by Corporation X" further restricts the name value domain to a very specific subset of people names: the names of Corporation X employees.

Therefore, Name of Person Employed by Corporation X has the value domain of all conceivable names limited to only those names of Corporation X employees. The number of all possible names valid for this data element is greatly reduced to a few specific ones.

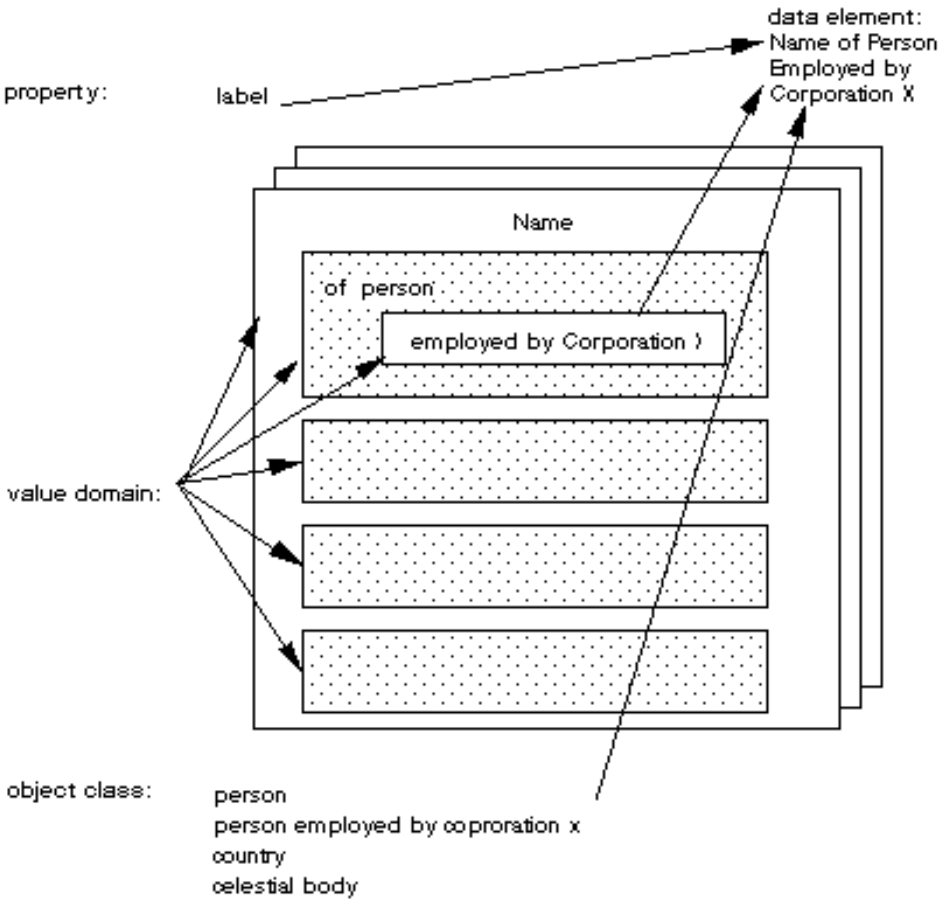


Figure A-6: Value Domain Restriction

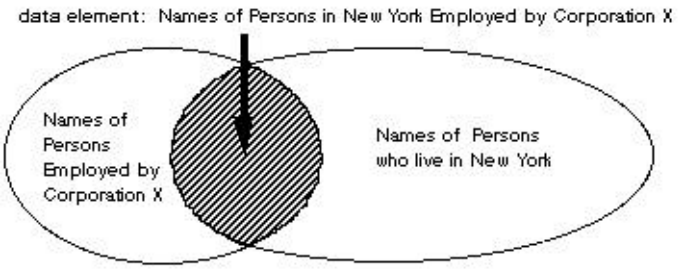


Figure A-7: Value Domain Intersection

Some enumerated domains can be considered an intersection of broader enumerated domains. Figure A-7 shows that, in some cases, value domains can overlap.

A.3.3.2 Reusable Domains

Some domains have the potential to be used by many data elements. For instance, the names of all the human eye colors may be applied to employees, contractors, spouses, and many other types of persons within a database.

A domain which is rigorously defined and specified may be reused. These reusable domains share the same value domain, representation, and property when used in a data element, for example employee eye color or contractor eye color. We call these reusable domains generic data elements (see Fig A-1).

A.4 Data Element Registry and Registration

A group appointed the responsibility for registering data elements is called a "Registration Authority." A Registration Authority:

- assigns a unique identifier
- assures that all required metadata attributes are documented
- assigns a status level depending upon the metadata quality and degree of integration.

The Registration Authority is appointed to be responsible for a *universe of discourse*. This universe may be as narrow as an organizational sub-unit within a company, or as wide as an entire industry or nation. The Registration Authority is responsible for the registration of any data element considered shareable within the universe of discourse for which it is appointed.

There will be a hierarchy of Registration Authorities, each responsible for a larger universe than its subordinate

Registration Authorities. Registration Authorities may be created at a higher level when there is a need to share data elements across two or more Registration Authorities at a lower level. The lower level then becomes the Submitting Organization for registration at the higher (with a broader universe of discourse) Registration Authority.
The Registration Authority documents the descriptions of data elements for which it is responsible in a Registry.
The Registration Authority is responsible for the integrity of the data elements in the Registry.

Annex B (informative)

Metadata and Data Elements

This informative annex discusses the connection between metadata and data elements.

B.1 What Is Metadata?

Commonly, metadata is defined as "data about data," but this is imprecise and leads to confusion. Metadata is the information and documentation which makes data understandable and shareable for users over time. Data remain useable, shareable, and understandable as long as the metadata remain accessible.

All organizations which produce data have an obligation to produce the metadata necessary to make the data understandable, both for internal and external users of the data. It is not sufficient for these users to have access to data without the information needed to understand or interpret the data.

Each organization which produces data also has an obligation to define what constitutes its metadata precisely. There are three reasons for this:

- A clear distinction must be made between data and metadata within a context (even though metadata is often treated as other data within information systems);
- The full extent of metadata must be determined, i.e. all the metadata necessary to describe the data fully; and
- General definitions of metadata fail because each organization's needs and uses for metadata differ so widely.

Because metadata is data (in a sense outside the needs of the organization), it is possible to store and retrieve metadata in a repository similar to the way data is stored and retrieved in a database. Efficient and effective use of data is made possible by the organized storage and use of metadata. Through on-line data dissemination, data and their metadata can be made available together. Data users will be able to get the information they need to use the data by looking at the metadata first.

Organizations produce different kinds of data, and therefore have different requirements for metadata. However, broad classification schemes for metadata have much in common across organizations. A commonly found scheme is the following:

- System - the physical and logical characteristics needed for computer programming and database management, includes information such as file locations, storage media, record layouts, database schema, data dictionaries, etc;
- Application - the information needed to understand or use the data, information such as definitions of terms, collection procedures and instruments, post data collection processing, etc;
- Administration - the costs, schedules, budgets, and related information associated with data collection programs and used by management for analysis.

As described above, metadata is data and can be organized formally through the use of data models and metamodels. A thorough understanding of all the metadata that any data user might want is a first step towards building these models. Then, the models will be the basis for building a repository of the metadata. The models also form the basis for classifying some metadata, thus they form part of a classification scheme.

Another view of metadata is gained by looking at the possible users of the data and metadata. Each type of user has

different needs, and some users' needs may not overlap. A programmer may only need to know system metadata, whereas an executive might only need to look at administration metadata. Data analysts typically have much broader needs, especially if they are external to the organization producing the data. In general, the different types of users, their needs (based on the types of questions they want to answer with the data), and their level of sophistication all have an impact on the metadata which has to be made available to them.

B.2 Data Elements, Metadata, Metamodels

The associated information about data elements is an integral part of any organization's metadata. Parts 2-6 of ISO/IEC 11179 describe the metadata which is required to be maintained for the data elements of an organization. These metadata will help all users understand and share an organization's data. Classification, basic attributes, definitions, naming, and registration are all important categories of information for the potential users of data.

Organizing and storing metadata in a repository requires building models. A metamodel of the information described in Parts 2-6 of ISO/IEC 11179 is necessary to capture that information in a registry or repository. It will not be easy to develop such a metamodel, but its development will greatly enhance the usefulness of an implementation of this standard.

Metadata about the data elements representing some data will rarely contain all the information necessary to understand those data. Much of the subject matter and administrative metadata about data and an organization is missing from the standard, and computer processing metadata is almost totally lacking, too. For instance, an organization collecting survey data will have much information about how the data were collected. This information, such as sample and questionnaire design, is often necessary to understanding the data. However, sample and questionnaire designs are not part of the metadata describing data elements (from ISO/IEC 11179).

Building a metamodel that is extensible will allow for additional metadata to be described. Subject matter, computer processing, and administrative areas which are beyond the scope of ISO/IEC 11179 need their own metamodels and data models developed. Extensibility will give the new models a place to fit into the basic data elements metamodel. In this way, a more complete model can be built.

Bibliography

Barr, Avron & Feigenbaum, Edward A., *The Handbook of Artificial Intelligence* (3 Volumes), William Kaufman, Inc., 1981.

Gosh, Sakti P., *Data Base Organization for Data Management*, Second Edition, New York: Academic Press, Inc., 1986.

Herman, G.T., "Theory of Algorithms", in *Encyclopedia of Computer Science and Engineering*, by Ralston, Anthony & Reilly, Jr., Edwin D., Second Edition, Van Nostrand Reinhold Company Inc., 1983.

ISO 639:1988, *Code for the representation of names of languages*

ISO 646, *ISO seven-bit coded character set for information interchange*

ISO 646:1983, *Information interchange - ISO 7-bit coded character set for information interchange*

ISO 2375, *Procedures for registration of escape sequence*

ISO 2788:1986, *Documentation - Guidelines for the establishment and development of monolingual thesauri*

ISO 3166:1988, *Code for the representation of names of countries ISO 4873, ISO eight-bit code for information interchange - structure and rules for implementation*

ISO 5985:1985 *Documentation - Guidelines for the establishment and development of multilingual thesauri*

ISO 6093, *Representation of numeric values in character strings*

ISO 6093:1985, *Information processing - Representation of numerical values in character strings for information interchange*

ISO 6862:1996, *Mathematical coded character set for bibliographic information interchange*

ISO 6937, *Coded character sets for text communication*

ISO 7498:1984, *Reference model of open systems interconnection*

ISO 8601:1988, *Data elements and interchange formats; - Representation of dates and times*

ISO 8824:1989, *Abstract Syntax Notation One*

ISO 8859:1992, *Eight-bit single-byte coded graphic characters sets*

ISO 9007:1987, *Information processing systems - Concepts and terminology for the conceptual schema and the information base*

ISO/IEC TR 9789:1993, *Information technology - Guidelines for the organization and representation of data elements for data interchange -- Coding methods and principles*

ISO/IEC 10027:1990, *Information technology - Information Resource Dictionary System (IRDS) Framework*

ISO/IEC 10032:1995, *Information technology - Reference model of data management*

ISO/IEC 10646-1:1993, *Universal Multiple-Octet Coded Character Set (UCS) - Part 1: Architecture and basic multilingual plane*

Manual for Data Administration, edited by Judith J. Newton and Daniel C. Wahl, National Institute of Standards and Technology, 1993.

Tasker, Dan, *Fourth Generation Data*, Prentice Hall, 1989.